

Semantics and Verification of Software

Lecture 18: Dataflow Analysis I (Introduction & Available Expressions Analysis)

Thomas Noll

Lehrstuhl für Informatik 2
(Software Modeling and Verification)

RWTH Aachen University

`noll@cs.rwth-aachen.de`

`http://www-i2.informatik.rwth-aachen.de/i2/svsw10/`

Summer Semester 2010

- 1 Preliminaries on Dataflow Analysis
- 2 An Example: Available Expressions Analysis

Dataflow Analysis: the Approach

- Traditional form of **program analysis**

Dataflow Analysis: the Approach

- Traditional form of **program analysis**
- Idea: describe how analysis information **flows** through program

Dataflow Analysis: the Approach

- Traditional form of **program analysis**
- Idea: describe how analysis information **flows** through program
- Distinctions:
 - direction of flow: **forward** vs. **backward** analyses
 - procedures: **interprocedural** vs. **intraprocedural** analyses
 - quantification over paths: **may** (**union**) vs. **must** (**intersection**) analyses
 - dependence on statement order: **flow-sensitive** vs. **flow-insensitive** analyses
 - distinction of procedure calls: **context-sensitive** vs. **context-insensitive** analyses

Labelled Programs

- Goal: **localization** of analysis information

Labelled Programs

- Goal: **localization** of analysis information
- Dataflow information will be associated with
 - **skip** statements
 - assignments
 - tests in conditionals (**if**) and loops (**while**)

These constructs will be called **blocks**.

Labelled Programs

- Goal: **localization** of analysis information
- Dataflow information will be associated with
 - **skip** statements
 - assignments
 - tests in conditionals (**if**) and loops (**while**)

These constructs will be called **blocks**.

- Assume set of **labels** L with meta variable $l \in L$
(usually $L = \mathbb{N}$)

Labelled Programs

- Goal: **localization** of analysis information
- Dataflow information will be associated with
 - **skip** statements
 - assignments
 - tests in conditionals (**if**) and loops (**while**)

These constructs will be called **blocks**.

- Assume set of **labels** L with meta variable $l \in L$
(usually $L = \mathbb{N}$)

Definition 18.1 (Labelled WHILE programs)

The **syntax of labelled WHILE programs** is defined by the following context-free grammar:

$$\begin{aligned} a &::= z \mid x \mid a_1 + a_2 \mid a_1 - a_2 \mid a_1 * a_2 \in AExp \\ b &::= t \mid a_1 = a_2 \mid a_1 > a_2 \mid \neg b \mid b_1 \wedge b_2 \mid b_1 \vee b_2 \in BExp \\ c &::= [\text{skip}]^l \mid [x := a]^l \mid c_1; c_2 \mid \\ &\quad \text{if } [b]^l \text{ then } c_1 \text{ else } c_2 \mid \text{while } [b]^l \text{ do } c \in Cmd \end{aligned}$$

Here all labels in a statement $c \in Cmd$ are assumed to be distinct.

Example 18.2

```
x := 6;  
y := 7;  
z := 0;  
while x > 0 do  
  x := x - 1;  
  v := y;  
  while v > 0 do  
    v := v - 1;  
    z := z + 1;
```

Example 18.2

```
[x := 6]1;  
[y := 7]2;  
[z := 0]3;  
while [x > 0]4 do  
  [x := x - 1]5;  
  [v := y]6;  
  while [v > 0]7 do  
    [v := v - 1]8;  
    [z := z + 1]9
```

Representing Control Flow I

Every (labelled) statement has a single entry (given by the initial label) and generally multiple exits (given by the final labels):

Definition 18.3 (Initial and final labels)

The mapping $\text{init} : \text{Cmd} \rightarrow L$ returns the **initial label** of a statement:

$$\begin{aligned}\text{init}([\text{skip}]^l) &:= l \\ \text{init}([x := a]^l) &:= l \\ \text{init}(c_1; c_2) &:= \text{init}(c_1) \\ \text{init}(\text{if } [b]^l \text{ then } c_1 \text{ else } c_2) &:= l \\ \text{init}(\text{while } [b]^l \text{ do } c) &:= l\end{aligned}$$

Representing Control Flow I

Every (labelled) statement has a single entry (given by the initial label) and generally multiple exits (given by the final labels):

Definition 18.3 (Initial and final labels)

The mapping $\text{init} : \text{Cmd} \rightarrow L$ returns the **initial label** of a statement:

$$\begin{aligned}\text{init}([\text{skip}]^l) &:= l \\ \text{init}([x := a]^l) &:= l \\ \text{init}(c_1; c_2) &:= \text{init}(c_1) \\ \text{init}(\text{if } [b]^l \text{ then } c_1 \text{ else } c_2) &:= l \\ \text{init}(\text{while } [b]^l \text{ do } c) &:= l\end{aligned}$$

The mapping $\text{final} : \text{Cmd} \rightarrow 2^L$ returns the set of **final labels** of a statement:

$$\begin{aligned}\text{final}([\text{skip}]^l) &:= \{l\} \\ \text{final}([x := a]^l) &:= \{l\} \\ \text{final}(c_1; c_2) &:= \text{final}(c_2) \\ \text{final}(\text{if } [b]^l \text{ then } c_1 \text{ else } c_2) &:= \text{final}(c_1) \cup \text{final}(c_2) \\ \text{final}(\text{while } [b]^l \text{ do } c) &:= \{l\}\end{aligned}$$

Definition 18.4 (Flow relation)

Given a statement $c \in \text{Cmd}$, the **(control) flow relation** $\text{flow}(c) \subseteq L \times L$ is defined by

$$\begin{aligned}\text{flow}([\text{skip}]^l) &:= \emptyset \\ \text{flow}([x := a]^l) &:= \emptyset \\ \text{flow}(c_1; c_2) &:= \text{flow}(c_1) \cup \text{flow}(c_2) \cup \\ &\quad \{(l, \text{init}(c_2)) \mid l \in \text{final}(c_1)\} \\ \text{flow}(\text{if } [b]^l \text{ then } c_1 \text{ else } c_2) &:= \text{flow}(c_1) \cup \text{flow}(c_2) \cup \\ &\quad \{(l, \text{init}(c_1)), (l, \text{init}(c_2))\} \\ \text{flow}(\text{while } [b]^l \text{ do } c) &:= \text{flow}(c) \cup \{(l, \text{init}(c))\} \cup \\ &\quad \{(l', l) \mid l' \in \text{final}(c)\}\end{aligned}$$

Example 18.5

```
 $c = [z := 1]^1;$   
  while  $[x > 0]^2$  do  
     $[z := z*y]^3;$   
     $[x := x-1]^4$ 
```

Example 18.5

```
 $c = [z := 1]^1;$   
   $\text{while } [x > 0]^2 \text{ do}$   
     $[z := z*y]^3;$   
     $[x := x-1]^4$ 
```

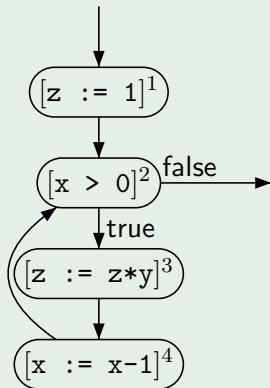
$\text{init}(c) = 1$
 $\text{final}(c) = \{2\}$
 $\text{flow}(c) = \{(1, 2), (2, 3), (3, 4), (4, 2)\}$

Example 18.5

```
c = [z := 1]1;  
  while [x > 0]2 do  
    [z := z*y]3;  
    [x := x-1]4
```

```
init(c) = 1  
final(c) = {2}  
flow(c) = {(1, 2), (2, 3), (3, 4), (4, 2)}
```

Visualization by
(control) flow graph:



Representing Control Flow IV

- To simplify the presentation we will often assume that the program $c \in Cmd$ under consideration has an **isolated entry**, meaning that

$$\{l \in L \mid (l, \text{init}(c)) \in \text{flow}(c)\} = \emptyset$$

(which is the case when c does not start with a **while** loop)

Representing Control Flow IV

- To simplify the presentation we will often assume that the program $c \in Cmd$ under consideration has an **isolated entry**, meaning that

$$\{l \in L \mid (l, \text{init}(c)) \in \text{flow}(c)\} = \emptyset$$

(which is the case when c does not start with a **while** loop)

- Similarly: $c \in Cmd$ has **isolated exits** if

$$\{l' \in L \mid (l, l') \in \text{flow}(c) \text{ for some } l \in \text{final}(c)\} = \emptyset$$

Representing Control Flow IV

- To simplify the presentation we will often assume that the program $c \in Cmd$ under consideration has an **isolated entry**, meaning that

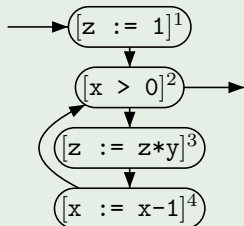
$$\{l \in L \mid (l, \text{init}(c)) \in \text{flow}(c)\} = \emptyset$$

(which is the case when c does not start with a **while** loop)

- Similarly: $c \in Cmd$ has **isolated exits** if

$$\{l' \in L \mid (l, l') \in \text{flow}(c) \text{ for some } l \in \text{final}(c)\} = \emptyset$$

Example 18.6



has an isolated entry but not isolated exits

- 1 Preliminaries on Dataflow Analysis
- 2 An Example: Available Expressions Analysis

Goal of the Analysis

Available Expressions Analysis

The goal of **Available Expressions Analysis** is to determine, for each program point, which (complex) expressions *must* have been computed, and not later modified, on all paths to the program point.

Available Expressions Analysis

The goal of **Available Expressions Analysis** is to determine, for each program point, which (complex) expressions *must* have been computed, and not later modified, on all paths to the program point.

- can be used to **avoid recomputations** of expressions
- only interesting for non-trivial (i.e., complex) arithmetic expressions

Goal of the Analysis

Available Expressions Analysis

The goal of **Available Expressions Analysis** is to determine, for each program point, which (complex) expressions *must* have been computed, and not later modified, on all paths to the program point.

- can be used to **avoid recomputations** of expressions
- only interesting for non-trivial (i.e., complex) arithmetic expressions

Example 18.7 (Available Expressions Analysis)

```
[x := a+b]1;  
[y := a*b]2;  
while [y > a+b]3 do  
  [a := a+1]4;  
  [x := a+b]5
```


Goal of the Analysis

Available Expressions Analysis

The goal of **Available Expressions Analysis** is to determine, for each program point, which (complex) expressions *must* have been computed, and not later modified, on all paths to the program point.

- can be used to **avoid recomputations** of expressions
- only interesting for non-trivial (i.e., complex) arithmetic expressions

Example 18.7 (Available Expressions Analysis)

```
[x := a+b]1;  
[y := a*b]2;  
while [y > a+b]3 do  
  [a := a+1]4;  
  [x := a+b]5
```

- **a+b** available at label 3

Goal of the Analysis

Available Expressions Analysis

The goal of **Available Expressions Analysis** is to determine, for each program point, which (complex) expressions *must* have been computed, and not later modified, on all paths to the program point.

- can be used to **avoid recomputations** of expressions
- only interesting for non-trivial (i.e., complex) arithmetic expressions

Example 18.7 (Available Expressions Analysis)

```
[x := a+b]1;  
[y := a*b]2;  
while [y > a+b]3 do  
  [a := a+1]4;  
  [x := a+b]5
```

- a+b available at label 3
- **a+b not available at label 5**

Goal of the Analysis

Available Expressions Analysis

The goal of **Available Expressions Analysis** is to determine, for each program point, which (complex) expressions *must* have been computed, and not later modified, on all paths to the program point.

- can be used to **avoid recomputations** of expressions
- only interesting for non-trivial (i.e., complex) arithmetic expressions

Example 18.7 (Available Expressions Analysis)

```
[x := a+b]1;  
[y := a*b]2;  
while [y > a+b]3 do  
  [a := a+1]4;  
  [x := a+b]5
```

- a+b available at label 3
- a+b not available at label 5
- possible optimization:
 while [y > **x**]³ do

- Given $c \in \text{Cmd}$, $L_c / \text{Block}_c / \text{AExp}_c$ denote the sets of all labels/blocks/complex arithmetic expressions occurring in c , respectively

Formalizing Available Expressions Analysis I

- Given $c \in Cmd$, $L_c/Block_c/AExp_c$ denote the sets of all labels/blocks/complex arithmetic expressions occurring in c , respectively
- An expression a is **killed** in a block B if any of the variables in a is modified in B

- Given $c \in \text{Cmd}$, $L_c / \text{Block}_c / AExp_c$ denote the sets of all labels/blocks/complex arithmetic expressions occurring in c , respectively
- An expression a is **killed** in a block B if any of the variables in a is modified in B
- Formally: $\text{kill}_{AE} : \text{Block}_c \rightarrow 2^{AExp_c}$ is defined by
$$\begin{aligned}\text{kill}_{AE}([\text{skip}]^l) &:= \emptyset \\ \text{kill}_{AE}([x := a]^l) &:= \{a' \in AExp_c \mid x \in FV(a')\} \\ \text{kill}_{AE}([b]^l) &:= \emptyset\end{aligned}$$

- Given $c \in Cmd$, $L_c/Block_c/AExp_c$ denote the sets of all labels/blocks/complex arithmetic expressions occurring in c , respectively
- An expression a is **killed** in a block B if any of the variables in a is modified in B
- Formally: $kill_{AE} : Block_c \rightarrow 2^{AExp_c}$ is defined by
$$\begin{aligned} kill_{AE}([skip]^l) &:= \emptyset \\ kill_{AE}([x := a]^l) &:= \{a' \in AExp_c \mid x \in FV(a')\} \\ kill_{AE}([b]^l) &:= \emptyset \end{aligned}$$
- An expression a is **generated** in a block B if it is evaluated in and none of its variables are modified by B

- Given $c \in Cmd$, $L_c/Block_c/AExp_c$ denote the sets of all labels/blocks/complex arithmetic expressions occurring in c , respectively
- An expression a is **killed** in a block B if any of the variables in a is modified in B
- Formally: $kill_{AE} : Block_c \rightarrow 2^{AExp_c}$ is defined by
$$\begin{aligned}kill_{AE}([skip]^l) &:= \emptyset \\kill_{AE}([x := a]^l) &:= \{a' \in AExp_c \mid x \in FV(a')\} \\kill_{AE}([b]^l) &:= \emptyset\end{aligned}$$
- An expression a is **generated** in a block B if it is evaluated in and none of its variables are modified by B
- Formally: $gen_{AE} : Block_c \rightarrow 2^{AExp_c}$ is defined by
$$\begin{aligned}gen_{AE}([skip]^l) &:= \emptyset \\gen_{AE}([x := a]^l) &:= \{a \mid x \notin FV(a)\} \\gen_{AE}([b]^l) &:= AExp_b\end{aligned}$$

Example 18.8 ($\text{kill}_{\text{AE}}/\text{gen}_{\text{AE}}$ functions)

```
 $c = [x := a+b]^1;$   
     $[y := a*b]^2;$   
    while  $[y > a+b]^3$  do  
         $[a := a+1]^4;$   
         $[x := a+b]^5$ 
```

Example 18.8 ($\text{kill}_{\text{AE}}/\text{gen}_{\text{AE}}$ functions)

```
 $c = [x := a+b]^1;$   
   $[y := a*b]^2;$   
  while  $[y > a+b]^3$  do  
     $[a := a+1]^4;$   
     $[x := a+b]^5$ 
```

- $AExp_c = \{a+b, a*b, a+1\}$

Example 18.8 ($\text{kill}_{\text{AE}}/\text{gen}_{\text{AE}}$ functions)

```
 $c = [x := a+b]^1;$   
 $[y := a*b]^2;$   
while  $[y > a+b]^3$  do  
   $[a := a+1]^4;$   
   $[x := a+b]^5$ 
```

- $AExp_c = \{a+b, a*b, a+1\}$

- | L_c | $\text{kill}_{\text{AE}}(B^l)$ | $\text{gen}_{\text{AE}}(B^l)$ |
|-------|--------------------------------|-------------------------------|
| 1 | $\emptyset$ | $\{a+b\}$ |
| 2 | $\emptyset$ | $\{a*b\}$ |
| 3 | $\emptyset$ | $\{a+b\}$ |
| 4 | $\{a+b, a*b, a+1\}$ | $\emptyset$ |
| 5 | $\emptyset$ | $\{a+b\}$ |

The Equation System I

- Analysis itself defined by setting up an **equation system**

The Equation System I

- Analysis itself defined by setting up an **equation system**
- For each $l \in L_c$, $AE_l \subseteq AExp_c$ represents the **set of available expressions at the entry of block B^l**

The Equation System I

- Analysis itself defined by setting up an **equation system**
- For each $l \in L_c$, $AE_l \subseteq AExp_c$ represents the **set of available expressions at the entry of block B^l**
- Formally, for $c \in Cmd$ with isolated entry:

$$AE_l = \begin{cases} \emptyset & \text{if } l = \text{init}(c) \\ \bigcap \{ \varphi_{l'}(AE_{l'}) \mid (l', l) \in \text{flow}(c) \} & \text{otherwise} \end{cases}$$

where $\varphi_{l'} : 2^{AExp_c} \rightarrow 2^{AExp_c}$ denotes the **transfer function** of block $B^{l'}$, given by

$$\varphi_{l'}(A) := (A \setminus \text{kill}_{AE}(B^{l'})) \cup \text{gen}_{AE}(B^{l'})$$

The Equation System I

- Analysis itself defined by setting up an **equation system**
- For each $l \in L_c$, $AE_l \subseteq AExp_c$ represents the **set of available expressions at the entry of block B^l**
- Formally, for $c \in Cmd$ with isolated entry:

$$AE_l = \begin{cases} \emptyset & \text{if } l = \text{init}(c) \\ \bigcap \{ \varphi_{l'}(AE_{l'}) \mid (l', l) \in \text{flow}(c) \} & \text{otherwise} \end{cases}$$

where $\varphi_{l'} : 2^{AExp_c} \rightarrow 2^{AExp_c}$ denotes the **transfer function** of block $B^{l'}$, given by

$$\varphi_{l'}(A) := (A \setminus \text{kill}_{AE}(B^{l'})) \cup \text{gen}_{AE}(B^{l'})$$

- Characterization of analysis:

forward: starts in $\text{init}(c)$ and proceeds downwards

must: $\bigcap$ in equation for AE_l

flow-sensitive: results depending on order of assignments

The Equation System I

- Analysis itself defined by setting up an **equation system**
- For each $l \in L_c$, $AE_l \subseteq AExp_c$ represents the **set of available expressions at the entry of block B^l**
- Formally, for $c \in Cmd$ with isolated entry:

$$AE_l = \begin{cases} \emptyset & \text{if } l = \text{init}(c) \\ \bigcap \{ \varphi_{l'}(AE_{l'}) \mid (l', l) \in \text{flow}(c) \} & \text{otherwise} \end{cases}$$

where $\varphi_{l'} : 2^{AExp_c} \rightarrow 2^{AExp_c}$ denotes the **transfer function** of block $B^{l'}$, given by

$$\varphi_{l'}(A) := (A \setminus \text{kill}_{AE}(B^{l'})) \cup \text{gen}_{AE}(B^{l'})$$

- Characterization of analysis:
 - forward**: starts in $\text{init}(c)$ and proceeds downwards
 - must**: $\bigcap$ in equation for AE_l
 - flow-sensitive**: results depending on order of assignments
- Later: solution **not necessarily unique**
 $\implies$ choose **greatest one**

The Equation System II

Reminder:

$$\begin{aligned} \text{AE}_l &= \begin{cases} \emptyset & \text{if } l = \text{init}(c) \\ \bigcap \{ \varphi_{l'}(\text{AE}_{l'}) \mid (l', l) \in \text{flow}(c) \} & \text{otherwise} \end{cases} \\ \varphi_{l'}(E) &= (E \setminus \text{kill}_{\text{AE}}(B^{l'})) \cup \text{gen}_{\text{AE}}(B^{l'}) \end{aligned}$$

The Equation System II

Reminder: $AE_l = \begin{cases} \emptyset & \text{if } l = \text{init}(c) \\ \bigcap \{ \varphi_{l'}(AE_{l'}) \mid (l', l) \in \text{flow}(c) \} & \text{otherwise} \end{cases}$
 $\varphi_{l'}(E) = (E \setminus \text{kill}_{AE}(B^{l'})) \cup \text{gen}_{AE}(B^{l'})$

Example 18.9 (AE equation system)

```
c = [x := a+b]1;  
    [y := a*b]2;  
    while [y > a+b]3 do  
        [a := a+1]4;  
        [x := a+b]5
```

The Equation System II

Reminder: $AE_l = \begin{cases} \emptyset & \text{if } l = \text{init}(c) \\ \bigcap \{ \varphi_{l'}(AE_{l'}) \mid (l', l) \in \text{flow}(c) \} & \text{otherwise} \end{cases}$
 $\varphi_{l'}(E) = (E \setminus \text{kill}_{AE}(B^{l'})) \cup \text{gen}_{AE}(B^{l'})$

Example 18.9 (AE equation system)

```
c = [x := a+b]1;  
    [y := a*b]2;  
    while [y > a+b]3 do  
        [a := a+1]4;  
        [x := a+b]5
```

$l \in L_c$	$\text{kill}_{AE}(B^l)$	$\text{gen}_{AE}(B^l)$
1	$\emptyset$	$\{a+b\}$
2	$\emptyset$	$\{a*b\}$
3	$\emptyset$	$\{a+b\}$
4	$\{a+b, a*b, a+1\}$	$\emptyset$
5	$\emptyset$	$\{a+b\}$

The Equation System II

Reminder: $AE_l = \begin{cases} \emptyset & \text{if } l = \text{init}(c) \\ \bigcap \{ \varphi_{l'}(AE_{l'}) \mid (l', l) \in \text{flow}(c) \} & \text{otherwise} \end{cases}$
 $\varphi_{l'}(E) = (E \setminus \text{kill}_{AE}(B^{l'})) \cup \text{gen}_{AE}(B^{l'})$

Example 18.9 (AE equation system)

```
c = [x := a+b]1;  
    [y := a*b]2;  
    while [y > a+b]3 do  
        [a := a+1]4;  
        [x := a+b]5
```

Equations:

$$AE_1 = \emptyset$$

$$AE_2 = \varphi_1(AE_1) = AE_1 \cup \{a+b\}$$

$$\begin{aligned} AE_3 &= \varphi_2(AE_2) \cap \varphi_5(AE_5) \\ &= (AE_2 \cup \{a*b\}) \cap (AE_5 \cup \{a+b\}) \end{aligned}$$

$$AE_4 = \varphi_3(AE_3) = AE_3 \cup \{a+b\}$$

$$AE_5 = \varphi_4(AE_4) = AE_4 \setminus \{a+b, a*b, a+1\}$$

$l \in L_c$	$\text{kill}_{AE}(B^l)$	$\text{gen}_{AE}(B^l)$
1	$\emptyset$	$\{a+b\}$
2	$\emptyset$	$\{a*b\}$
3	$\emptyset$	$\{a+b\}$
4	$\{a+b, a*b, a+1\}$	$\emptyset$
5	$\emptyset$	$\{a+b\}$

The Equation System II

Reminder: $AE_l = \begin{cases} \emptyset & \text{if } l = \text{init}(c) \\ \bigcap \{ \varphi_{l'}(AE_{l'}) \mid (l', l) \in \text{flow}(c) \} & \text{otherwise} \end{cases}$
 $\varphi_{l'}(E) = (E \setminus \text{kill}_{AE}(B^{l'})) \cup \text{gen}_{AE}(B^{l'})$

Example 18.9 (AE equation system)

```
c = [x := a+b]1;  
    [y := a*b]2;  
    while [y > a+b]3 do  
        [a := a+1]4;  
        [x := a+b]5
```

$l \in L_c$	$\text{kill}_{AE}(B^l)$	$\text{gen}_{AE}(B^l)$
1	$\emptyset$	$\{a+b\}$
2	$\emptyset$	$\{a*b\}$
3	$\emptyset$	$\{a+b\}$
4	$\{a+b, a*b, a+1\}$	$\emptyset$
5	$\emptyset$	$\{a+b\}$

Equations:

$$AE_1 = \emptyset$$

$$AE_2 = \varphi_1(AE_1) = AE_1 \cup \{a+b\}$$

$$\begin{aligned} AE_3 &= \varphi_2(AE_2) \cap \varphi_5(AE_5) \\ &= (AE_2 \cup \{a*b\}) \cap (AE_5 \cup \{a+b\}) \end{aligned}$$

$$AE_4 = \varphi_3(AE_3) = AE_3 \cup \{a+b\}$$

$$AE_5 = \varphi_4(AE_4) = AE_4 \setminus \{a+b, a*b, a+1\}$$

Solution: $AE_1 = \emptyset$
 $AE_2 = \{a+b\}$
 $AE_3 = \{a+b\}$
 $AE_4 = \{a+b\}$
 $AE_5 = \emptyset$